THE GENAI DIVIDE · WHAT THE 95% ACTUALLY MEASURES

The Number Is Roughly Right. The Reading Is Wrong.

Why MIT's 95% measures a comprehension gap, not a technology failure — and why comprehension gaps and calculation gaps demand opposite responses

AUDIENCE
CIOs, VPs Ops, Mining Directors

FOCUS
Reading the MIT 95% by failure type

REGION
LATAM heavy industry · Chile

NTT DATA · Technology Consulting · 2026 · biztalksnttdata.com

Contents

01

MIT DID NOT MEASURE WHAT THE HEADLINE SAYS

MIT's "95% without ROI" is not a measurement of AI failure. It is a measurement of distance: the gap between what a company believed it bought and what the technology bills for before it pays. The number is roughly right. The standard reading of it is wrong, and the difference between the two is worth several quarters of your AI budget.

The figure comes from *The GenAI Divide: State of AI in Business 2025*, published in August 2025 by Project NANDA (Networked Agents and Decentralized AI), a research group at the MIT Media Lab. Before building anything on top of it, read what it is. The study rests on fifty-two structured executive interviews, 153 survey responses gathered from senior leaders at industry conferences, and a review of more than 300 publicly disclosed AI initiatives whose synthesis method was never published. "Measurable ROI" (return on investment) was defined by each respondent, not by a common standard, and the 95% refers to organizations reporting no measurable profit-and-loss impact — not to pilots counted as isolated units. The authors flagged all of this themselves: the report carries the disclaimer "preliminary findings, anonymized for confidentiality," and describes its conclusions as directionally accurate based on individual interviews rather than official company reporting. It is a snapshot taken early on a capability curve that moves in quarters, not decades.

That distinction is the whole game. A directional signal tells you which way the river runs. A precise measurement tells you how deep it is at a named point. The viral retelling of the MIT number treats a directional signal as a precise measurement, then draws a precise conclusion from it: *95% of AI pilots fail, therefore AI does not work.* The premise is soft. The conclusion is hard. The reasoning in between is broken.

A company building a thesis on the MIT number has to hold two things at once. The signal is real: most enterprise GenAI (generative AI) spend in 2025 returned nothing the finance function could see. And the signal is coarse: it does not tell you, for any single portfolio, *why* the return was invisible. The "why" is not in the report. The "why" is in your own operation, and there are at least two of them, with opposite remedies.

This piece argues one claim and defends it: the 95% measures a comprehension gap, not a verdict on the technology. A zero on the ROI line is a symptom. The same symptom is produced by two different conditions that look identical to a CFO and require opposite responses. Naming which one you have is the deliverable. Everything else is commentary.

02

THE NUMBER ON THE BOARD SLIDE, TRANSLATED TO A COPPER PORTFOLIO

The MIT statistic arrives in a boardroom as a single sentence on a forwarded article, usually from a director who read it on a flight. By the time it lands in front of the executive who signed the AI portfolio, it has been compressed into one question: *should we pause?*

That question cannot be answered at the altitude it is asked. "Should we pause the AI program" assumes there is one program with one health status. A real portfolio in a copper producer is not one program. It is a major concentrator running a short-interval control optimizer on the grinding circuit, a planning office piloting a maintenance copilot, a metallurgical-accounting assistant a superintendent built without telling anyone, and three vendor relationships at three different stages of disappointment. Each of those has its own answer to "is it working." The board sees their average. The average is the least useful number in the room.

Consider what the average hides. A polymetallic operation in southern Peru running three AI vendors and one in-house data team can have one initiative producing real value, one producing the appearance of value, and one producing nothing, and report all three through a single "AI portfolio ROI" line that nets to roughly zero. The zero is arithmetically correct. As a decision input it is worthless, because the three initiatives need three different decisions: press, re-measure, stop.

This is the MIT 95% rendered at the resolution an operator actually works at. The report flattens millions of these portfolios into one number. The standard reading then flattens your portfolio the same way, and invites you to make one decision for initiatives that need three. The flattening is the error. The board's instinct to ask "pause or press" is the flattening in action.

The work, then, is to recover the distribution behind the average. Not the industry's distribution: yours. What fraction of your portfolio is producing value the ledger cannot see? What fraction is producing the appearance of value with no baseline behind it? What fraction is genuinely returning nothing, and of that fraction, how much is fixable with better information versus fixable only by changing who decides what? Those are answerable questions. The MIT number is not one of them. It is the prompt to start asking them.

03

THE YEAR THREE PILOTS CLOSED WITH ONE REPORTED RETURN

In the AI-portfolio reviews we have run with copper producers across LATAM, the pattern repeats with a regularity that no longer surprises us. One of those reviews captures it whole. It was a Thursday in late March, the closeout of the fiscal year, when the integrated planning office of a top-five regional mining group presented its three-year AI portfolio to the board. The CIO who had signed that portfolio had read the MIT statistic the week before, on a forward from a board member, and had come to the review prepared to recommend a pause. By the end of the meeting the recommendation had changed, because the review — run with our team in the room — surfaced something the portfolio summary had buried.

Three pilots had run over three years against a total spend of roughly USD 2.6 million.

The first was a short-interval control optimizer on the grinding circuit of a 95,000-tonne-per-day concentrator. It reported the only hard ROI number in the portfolio: a claimed 1.8% throughput gain, booked at an estimated USD 9.2 million per year. That number was the headline of the deck. It was also the only number anyone questioned, because two other things had changed in the same two quarters the optimizer went live: a scheduled mill reline that restored mechanical performance, and a new ore-blend recipe from the metallurgy group. Three causes, one moved indicator, and the AI pilot had claimed the entire delta.

The second was a maintenance-planning copilot. Its success metric was adoption: 80% of planners used it at least weekly. The deck reported the 80% as impact. No one had measured planning-cycle time before the copilot arrived, so there was no baseline against which “weekly use” could be converted into hours saved or downtime avoided. The activity was real. Whether it was a result, no one could say, because the question had not been instrumented before deployment.

The third pilot was not in the AI portfolio at all. A metallurgical-accounting superintendent, working with a quota of model credits and no steering committee, had built an assistant that pulled month-end production figures from the plant historian and the LIMS (laboratory information management system) and drafted the reconciliation that closed the metallurgical accounts. The close had taken six days for the two prior years. It now took two. Three analysts recovered roughly four days each month. The value was modest and real, perhaps USD 180,000 a year in recovered analyst time plus a faster, cleaner close. It had been booked under “finance process improvement.” It was never tagged as AI, so it never reached the AI ledger.

Read the portfolio as the board first saw it: one large ROI number under question, a pile of adoption metrics, and a total spend of USD 2.6 million. Read it as the review left it: one claimed return that three causes could explain, one pilot with no baseline, and one genuine return that the AI ledger could not see because it had been filed somewhere else. The portfolio’s apparent failure and the portfolio’s one real success were both products of how the numbers were filed, not of the technology.

The CIO did not pause the program. The CIO paused two pilots, kept the reconciliation assistant and asked finance to re-tag it, and sent the optimizer back to be re-measured

against the reline and the ore blend. That is not a story about whether AI works in mining. The technology worked in the one place it was pointed at a sharp problem with a measured baseline. It is a story about a number that meant the opposite of what it appeared to mean.

04

WHAT THE BUYER THOUGHT THEY PURCHASED, AND WHAT THE TECHNOLOGY BILLS FOR

The grinding optimizer, the planning copilot, and the reconciliation assistant failed and succeeded for the same underlying reason, expressed three ways. Each was governed by a contract that almost no one writes down, and the gap inside that contract is what the MIT 95% measures.

There are four parties to the contract every time an organization buys GenAI, and each signs a different version of it.

```
        THE UNWRITTEN COMPREHENSION CONTRACT

VENDOR SELLS     "Intelligence as a service. Connect your
                  data, ask, receive insight."

BUYER BUYS       "A productivity tool that pays back in
                  twelve months on the line we can see."

PLANT RECEIVES   "A probabilistic text engine with no
                  memory, no authority to decide, and no
                  wiring into the systems where work runs."

FINANCE MEASURES "Did the line in the income statement
                   move, and can we attribute the move."

 The gap between any two of these versions is a slice
 of the MIT 95%.
```

Figure 1 · The four versions of the unwritten contract. The 95% is the distance between what was sold, bought, received, and measured — not a property of the model.

The vendor sold a capability the technology does not yet have on its own. The buyer purchased a payback the budget never funded the scaffolding to produce. The plant received a tool that suggests but cannot enforce and remembers nothing between sessions without an architecture no one priced. Finance measured an income-statement line that none of the other three parties were actually moving. When the measured line reads zero, every party is surprised, and each blames a different one of the other three.

This is why the MIT 95% is a comprehension bubble, not a financial bubble. The capital was spent in full. The value was real in some pockets and absent in others. What inflated was not an asset price. What inflated was the distance between what four parties believed they had agreed to. A financial bubble deflates when a price collapses. A comprehension bubble deflates when the buyer reads the actual product specification and reconciles it against the budget, the operating chain, and the measurement frame.

The contract tightens when the four roles converge. The reconciliation assistant worked because one superintendent held all four: he sold himself nothing, bought a tool for a problem he had lived with, deployed it where the data already sat, and measured it against a six-day baseline he had watched for two years. The optimizer and the copilot loosened the contract across four different functions, none of which had read the others' version. The number that results is not a measurement of the model. It is a measurement of how far apart the four signatures sat.

05

WHY THE WALL STREET ARGUMENT DOESN'T REACH YOUR PLANT

A reader can collapse this argument into a different one, and a hostile reader will. The adjacent claim runs: AI is a bubble, valuations are detached from fundamentals, the correction is coming, so the prudent operator waits. That is the financial-bubble thesis. It is a real argument. It is not this one, and the distinction matters because either thesis can be true without the other.

The financial-bubble thesis is a claim about market prices: that what investors pay for AI exposure exceeds what the underlying businesses will earn. Defending it requires evidence this piece does not have and takes no position on: capital-flow data, valuation multiples, the cash conversion of the model vendors. Whether that thesis is true has no bearing on the argument here.

The comprehension-bubble thesis is a claim about buyer understanding: that the distance between what enterprises think they bought and what the technology requires to deliver value is large, and that this distance, not the technology's capability, accounts for most of the MIT 95%. It can be substantiated from inside a single portfolio, with the kind of review the planning office ran. It needs no market data at all.

Here is the sentence that separates them. *The comprehension bubble would still explain your zero-ROI line even if AI equities were fairly priced, and a financial-bubble correction would not, on its own, close the comprehension gap in your operation.* Suppose the market is rationally valued and AI is a sound long-term investment. Your portfolio can still read zero, because your four-party contract is loose. Suppose the market is wildly overvalued and a correction wipes out half the vendors. Your reconciliation assistant still saves four analyst-days a month, because its value never depended on a valuation.

The two theses are frequently fused in board conversation, and the fusion is expensive. It ties a defensible claim about your own operating reality to an undefended claim about market timing, and invites a skeptic to dismiss both by puncturing the one you were never

making. Keep them apart. Take no position on the financial bubble. You do not need one. The comprehension bubble is sitting in your last fiscal review, and you can act on it Monday without a single data point about anyone's market capitalization.

06

A DIRECTIONAL SIGNAL, READ AS A VERDICT

The MIT number went viral for a reason worth naming, because the reason is also the trap. A precise-sounding statistic that confirms a suspicion travels faster than a careful one that complicates it. "95% of AI pilots fail" confirmed what a tired executive already half-believed after a year of disappointing demos. It gave permission to stop. Permission travels.

The methodological critics arrived within days, not months. The Marketing AI Institute published its rebuttal on August 26, 2025 — eight days after the report's press launch — and put the problem precisely: the retelling dropped every qualifier the authors kept, and a qualifier-free statistic is a different object from the one the researchers published. The success bar itself was unusually strict: deployment beyond pilot with measurable ROI within six months, a definition that files efficiency gains and cost reduction under failure. The same week, Wharton professor Kevin Werbach issued the sharpest academic challenge on record: if Project NANDA stands behind the claims, it should release the full supporting data; if not, it should retract the report. As of mid-2026, NANDA has done neither. The researchers measured a tendency and said so. The internet reported a law, and the authors let the law circulate.

There is a second-order problem underneath even the careful reading, and it sharpens the case rather than weakening it. ROI may be the wrong instrument for this measurement in the first place. UC Berkeley's executive-education group argued in September 2025 that return on investment, calibrated for capital projects with predictable payback curves, mis-measures a capability that improves on a steep curve and delivers value in forms the income statement is slow to recognize. If the metric itself is miscalibrated for the asset, then a portfolio of zeros tells you as much about the ruler as about the thing being measured.

Hold the three observations together. The study is directional, not precise, by its own authors' framing. The retelling stripped the qualifiers that made it honest, and the unanswered challenge to the underlying data still stands. And the metric the retelling leans on may be the wrong metric for the asset. None of this rescues a genuinely failed AI program. All of it disqualifies the lazy inference that 95% means the technology does not work.

Naming these limits is not a hedge. It is the credibility move that separates a serious analysis from a slide that pattern-matched on a viral number. An executive who walks into a board meeting and says "the MIT study is directional, here is what it can and cannot tell us about our portfolio" has already won the argument against the director who read only the headline. The 90-second answer to "why does MIT say 95% fail" is this: it does not say that, it says something narrower and more useful, and here is how we read it inside our own operation.

07

THE EXECUTIVE WHO CANNOT SEE THE COST, AND THE ONE WHO HAS ALREADY PRICED IT

The zero on the ROI line is produced by two conditions that look identical to finance and are opposite at the level of remedy. Telling them apart is the central act of the diagnosis, because the wrong remedy applied to the right condition is one of the most reliable ways to manufacture another zero.

The first condition is a **failure of comprehension**. The executive who owns the initiative does not yet see the gap between what was bought and what the technology requires. The operating-model cost of capturing value, the decision rights to be moved, the data to be made reachable, the baseline to be measured, was never made visible, so it was never funded. This condition is contingent and corrigible. It dissolves when the cost is surfaced. Show the owner the four-party contract, name the scaffolding the budget skipped, and the owner can re-scope, re-measure, and continue. The disease is downstream of missing information, and information is the cure. This executive responds to evidence and argument.

The second condition is a **failure of calculation**. The executive sees the gap perfectly well. The decision rights are fragmented across three functions, the data sits in a function that will not share it, accountability for acting on the output is undefined, and the executive has already done the arithmetic: redesigning the operating model to fix this costs more, in political capital and disruption, than the ROI the AI program would have returned. So the program is allowed to fail quietly. This condition is structural, rooted in incentive and power, not in a blind spot. It does not respond to evidence, because the evidence is not missing. The executive already has it and has rationally chosen the lost ROI over the cost of redesign. This condition responds only to changed incentives or changed authority.

These are not two points on a spectrum. They are opposite diagnoses with opposite cures, and the cure for one is inert against the other.

Table 1 The diagnostic key. The same zero-ROI symptom; opposite conditions; opposite first moves.

Signal	Failure of comprehension	Failure of calculation
Does the owner see the operating-model cost?	No — it was never made visible	Yes — and has priced it against the lost ROI
What the owner needs	Information: the cost, named and quantified	Incentive or authority: a reason to pay the redesign cost
Response to a clear briefing	Re-scopes and presses forward	Nods, agrees, changes nothing
Ask “who decides whether to act on the output?”	A name, currently overcautious	Silence, or three functions with veto power

Correct first move	Press, with the contract rewritten	Stop the AI program; fix the operating chain or the incentive first

The reason this distinction is rarely drawn is that it is uncomfortable. A comprehension failure is a story about a gap in understanding, which a good briefing can close and which embarrasses no one permanently. A calculation failure is a story about an operating model whose incentives make the rational executive choose failure, which implicates the structure of the organization and the people who built it. Boards reach for the comprehension remedy because it is the cheap one, and apply it to calculation failures, where it does nothing but produce another year of pilots and another zero. The cheap remedy applied to the structural condition explains most of the MIT 95%.

08

WHAT A REPORTED WIN CONCEALS

If the 95% is not one population, neither is the 5%. The cohort that reports apparent ROI is itself a mix, and a technically literate reader should be as skeptical of a reported win as of a reported loss. Three patterns produce results that look like success without being it, and each has a single question that exposes it.

The first is **activity dressed as result**. An adoption metric is reported as impact: 80% of planners use the copilot weekly, therefore the copilot succeeded. The move conflates usage with value and survives only because no one measured the outcome before deployment. The detection question is *what was the baseline?* If there is no pre-deployment measurement of the thing the tool was supposed to improve, the reported win is an activity count wearing the costume of a result. The planning copilot in the signature story is this pattern exactly.

The second is **anecdote never multiplied**. One attractive case is retold until it stands in for the portfolio. The superintendent's reconciliation assistant is real and valuable, and a vendor or an internal champion will happily show it on every slide while the eleven other initiatives that returned nothing stay off the deck. The case is true. The generalization from it is false. The detection question is *what is the portfolio performance?* A single success proves the technology can work; it says nothing about whether your portfolio, on average, does. Demand the denominator, not the highlight reel.

The third is **indicator misattributed**. A metric genuinely moved, and the AI initiative claimed the credit, but several causes changed in the same window. The grinding optimizer's 1.8% throughput gain is this pattern: the mill reline and the ore-blend change are sitting in the same two quarters, each a candidate to explain the delta. The detection question is *what else changed in the same period?* If two or three plausible causes overlap the AI deployment and no one isolated them, the attribution is a guess dressed as a measurement.

These three questions convert the reader from a consumer of the 95% into an investigator of the 5%. *What was the baseline. What is the portfolio performance. What else changed in*

the same period. Run them against any AI win presented to you — your own, or a peer's story from the last industry dinner. A meaningful share of the 5% will reveal itself as the same comprehension gap from the other direction: value claimed without the instrumentation to defend it.

09

WHAT TO ASK BEFORE WEDNESDAY'S PORTFOLIO REVIEW

The diagnosis is not theoretical. It is a test you can run on a single initiative before your next portfolio review, with a binary outcome: the reported return is either defensible or it is not, and the failure type is either comprehension or calculation. Pick one stalled or one celebrated initiative and run five steps in order.

Step one. Name the baseline that was measured before deployment. If there is no pre-deployment number for the outcome the tool was meant to improve, stop: the reported result is unproven, and the condition is comprehension (no one made the measurement cost visible). If a baseline exists, continue.

Step two. Establish whether the reported result is portfolio-level or a single case. If it is one anecdote standing in for the program, the portfolio performance is unproven regardless of how good the case is. If it is a portfolio number, continue.

Step three. List everything else that changed in the same window: mechanical work, recipe changes, market moves, a new supervisor. If two or more plausible causes overlap the deployment and were not isolated, the attribution is unproven. If the cause is isolated, continue.

Step four. Ask the owner to name, in one sentence, the operating-model change required to capture the value: which decision moves, which data opens, who becomes accountable. If the owner cannot name it, the condition is comprehension: brief them, fund the scaffolding, press. If the owner names it precisely and then explains why it has not happened, continue.

Step five. Ask why the named change has not been made. If the answer is "no one showed us it was needed," it is comprehension after all. If the answer is "we know, but the cost of moving those decision rights exceeds what we'd get back," the condition is calculation: stop treating it as an AI problem, because no briefing will move it. It moves only when the incentive or the authority changes.

Five steps, one initiative, under an hour. The outcome is binary at each gate and the path through the gates classifies the failure. An initiative that passes all five gates has a defensible return. One that fails at step one, two, or three has an unproven return and a comprehension gap. One that reaches step five and stops there has a calculation gap, and the most expensive thing you can do is hand it another year of pilots.

If you cannot run this test on your own portfolio, you do not yet know what your AI program is doing, and neither does your board. That is the real finding the MIT number should provoke.

10

THE REMEDIES THAT TREAT THE WRONG DISEASE

Once a classification is in hand, the failure modes shift from "we did the wrong project" to "we applied the wrong remedy to the right diagnosis." Three of these are common enough in LATAM heavy industry to name.

The first is prescribing architecture for an operating-model question. A calculation failure surfaces, and the response is a lakehouse, an agent mesh, a retrieval layer, a new platform. The architecture is sold as the answer. It is not, because the condition was never technical. The decision rights are still fragmented after the platform ships, the data-owning function still will not share, and the executive who already priced the redesign and declined it now has a more expensive program returning the same zero. Architecture cannot buy authority. If the diagnosis is calculation, the first move is to change who decides or what they are rewarded for, and only then to build.

The second is treating the 95% as homogeneous and pausing everything. The board reads the number, applies one decision to a portfolio that needs three, and stops the comprehension-gap initiatives that were one scope reset from compounding alongside the calculation-gap initiatives that genuinely should stop. The cost of stopping a comprehension-gap program is not the budget written off; it is the organizational learning discarded, the workforce that now carries antibodies against the next AI initiative, and a restart from zero against a vendor landscape two generations further along. Slice the portfolio before you pause it. The average is not a decision.

The third is filing the operating-model cost as a soft factor. "Culture," "change management," "ways of working": the language demotes a hard cost line into a workshop. The persona reading this knows that decision rights, data ownership, and accountability are not soft. They are the determinants of whether any AI output gets acted on, and they belong in the business case as a quantified cost of capture, not in an appendix about adoption. When the operating-model change is priced as a line item, the calculation failure becomes visible as what it is: a rational choice the executive made because no one ever put the redesign cost and the foregone ROI on the same page.

The pattern across all three is identical. The remedy is chosen for the disease the management literature has trained the reflex for, not the disease the diagnosis found. Treat the condition you have. The cost of treating the other one is the next year's zero.

11

WHAT TO CONFIRM IN YOUR OWN PORTFOLIO THIS QUARTER

Four verifications, ordered from cheapest to most consequential. Each closes off a way the MIT average lies about your firm.

One. Run the five-step test from Section 09 against every active AI initiative, one at a time, and write the result down: defensible or unproven, comprehension or calculation. The cost is roughly an hour per initiative. The output is a classification of your entire portfolio, in writing, with evidence, which is precisely the distribution the MIT average withholds. By the next board cycle you can replace "AI portfolio ROI = zero" with a sentence that means something.

Two. For every initiative classified as comprehension, rewrite the four-party contract on a single page with the buyer, the deployer, and one line user in the room: what the tool does, what it does not, what counts as success, who acts on the output. A half-day each. The signed page is the cure for a comprehension gap, and it is cheap.

Three. For every initiative classified as calculation, take it out of the AI conversation entirely and put it in front of the executive who owns the operating chain, framed as a cost question: the redesign cost on one side, the foregone return on the other, on the same page. The AI program does not restart until that page has been read and a decision made. The honesty of that escalation is the test. An AI program must never be the cover story for an operating-model the organization has declined to fix.

Four. Audit the ledger for value booked under the wrong name. For every initiative reading zero, ask whether an improvement showed up elsewhere in the profit-and-loss statement: a faster close, less downtime, recovered analyst hours, and whether the improvement tracks the deployment timeline. Where it does, re-tag it prospectively and explain the re-tag at the next review. Expect finance to resist reopening a close it has already signed — which is why the re-tag goes forward, not backward. This is the one verification that can revise your reported number upward this quarter, and the reconciliation assistant in the signature story is the proof that the value is often already there, filed under something else.

Four verifications, one quarter. At the end you will know which initiatives to press, which to re-measure, which to stop, and which to rebook against the value line they were always producing. You will govern your AI portfolio by its distribution instead of by an average you could not interpret, which is the only honest way to answer the question the board actually asked.

The MIT 95% is a real number measuring an unreal thing: the average of two opposite conditions that produce one symptom. It tells you AI returned nothing you could see. It does not tell you whether the cure is a better briefing or a different operating model — and those are opposite. The number is the prompt. The diagnosis is the deliverable.

12

KEY INSIGHTS

The claims of this paper each stand on their own — one block per insight, a memory aid.

1 MIT's 95% is not a measurement of AI failure. It measures the distance between what a company believed it bought and what the technology requires before it pays — a comprehension gap, not a technology verdict.

2 Read the study before building on it: 52 interviews, 153 conference surveys, 300+ public initiatives with an unpublished synthesis method, ROI defined by each respondent. The authors themselves call it preliminary and directional. Directional is how it must be used.

3 The methodological challenges to the study — including a public call from Wharton to release the data or retract — went unanswered through mid-2026. A serious reader cites the number with its limits attached, or not at all.

4 Your portfolio average is the least useful number in the room. One zero-ROI line can hide three initiatives needing three opposite decisions: press, re-measure, stop. The average is not a decision.

5 Every GenAI purchase signs four versions of an unwritten contract — what the vendor sells, the buyer buys, the plant receives, and finance measures. The 95% is the distance between the four signatures.

6 The comprehension bubble is not the financial bubble, and either can be true without the other: your zero would persist even if AI equities were fairly priced, and a market correction would not close your operating gap. Defend the first; take no position on the second.

7 A failure of comprehension responds to information: surface the operating-model cost and the owner re-scopes and presses. A failure of calculation responds only to changed incentives or authority — the owner already saw the cost and rationally declined it.

8 The cheap remedy applied to the structural condition — briefing an executive who has already done the math — explains most of the MIT 95%.

9 The 5% is not one population either. Three patterns fake success: activity dressed as result, anecdote never multiplied, indicator misattributed. Three questions expose them: what was the baseline, what is the portfolio performance, what else changed in the same period.

10 Five gates, one hour, one initiative: baseline → portfolio-level → confounds isolated → operating-model change named → reason it has not happened. The path through the gates classifies the failure and dictates the first move.

11 Architecture cannot buy authority. A lakehouse shipped into a calculation failure is a more expensive program returning the same zero. Change who decides or what they are rewarded for, then build.

12 Audit the ledger for value booked under the wrong name — a faster close, recovered analyst hours, less downtime tracking the deployment timeline. It is the one verification that can revise your number upward this quarter.

13

SOURCES

The public claims in this paper, with their sources. The signature story and portfolio figures are anonymized composites drawn from LATAM natural-resources engagements; they illustrate the diagnostic, not a named client.

- MIT Project NANDA (Networked Agents and Decentralized AI), *The GenAI Divide: State of AI in Business 2025*, August 2025 — methodology, the 95% finding, and the authors' own "preliminary findings" and "directionally accurate" caveats — MIT Project NANDA.
- Press launch and CFO-audience coverage of the report (August 18, 2025) — Fortune.
- Mike Kaput, "That Viral MIT Study Claiming 95% of AI Pilots Fail? Don't Believe the Hype," Marketing AI Institute, August 26, 2025 — the qualifier-stripping critique and the six-month success-bar objection — Marketing AI Institute.
- "Why We Don't Believe MIT NANDA's Weird AI Study," Futuriom, August 26, 2025 — methodological objections, including Wharton professor Kevin Werbach's call to release the full supporting data or retract — Futuriom.
- David Gallacher, "Beyond ROI: Are We Using the Wrong Metric in Measuring AI Success?," UC Berkeley Executive Education, September 2025 — the argument that ROI itself is miscalibrated for the GenAI capability curve — UC Berkeley Executive Education.

NTT DATA · Technology Consulting · 2026 · biztalksnttdata.com

For more on this topic, contact: ***Jaime Rebolledo****, Director Value Proposition Mining, NTT DATA* · jaime.rebolledocampos@nttdata.com.